

Report on the Alfred P. Sloan Foundation Workshop

on

Protistan Barcoding, Reference Material and Cultures

held

November 6 and 7, 2006

at the

Portland Harbor Hotel, Portland, ME USA

by

Andersen, Robert A., Beakes, Gordon, Blackburn, Susan, Brand, Jerry, Cochrane, Guy, Draisma, Stefano, Duszynski, Donald, Edvardsen, Bente, Federhen, Scot, Friedl, Thomas, Hajibibael, Mehrdad, Hoef-Emden, Kerstin, Ishida, Ken, Kaczmarska, Irena, Karpov, Sergey, Kawachi, Masanobu, Kawai, Hiroshi, Küpper, Frithjof, Leander, Brian, Levesque, Andre, Lewis, Louise, Lynn, Denis, Moestrup, Øjvind, O'Kelly, Charles, Patterson, David, Phang, Siew Moi, Porter, David, Potapova, Marina, Probert, Ian, Pröschold, Thomas, Saunders, Gary, Schindel, David, Silva, Paul, Spiegel, Fred, Stern, Rowena, Theriot, Edward, Vargas, Colombar de, Vaultot, Daniel, and Williamson, Phil.

Table of Contents

I. Introduction.....	3
II. Summary of Workshop Conclusions.....	4
III. Background Information on Protists.....	5
IV. The Workshop.....	6
V. Problems and Recommended Solutions.....	7
Problem 1: How do we connect DNA from a culture to the correct name?.....	7
Problem 2. Which culture will be the “official” culture for anchoring the name?.....	8
Problem 3. How to connect names, cultures, barcodes, information for each species?.....	9
Problem 4. How do we implement the barcode/culture/name process?.....	10
VI. Appendices.....	9
Appendix A. Protist Diversity from Corliss 1982.....	11
Appendix B. Protist Species, Type Material, and Cultures from the Workshop.....	13
Appendix C. Workshop Participants.....	15
Appendix D. Agenda for the Workshop.....	17

I. Introduction

The identity of organisms has always been important for humans, whether for distinguishing food items for a cave man or recognizing invasive species for a present-day governmental official. Identifying organisms is a dynamic scientific field, often referred to as taxonomy. The practice of taxonomy is changing rapidly due to two technological breakthroughs. First, DNA is complementing traditional morphology for identifying organisms. Short sequences of DNA, termed barcodes, provide a precise means for species identification. Rapid, accurate identification of organisms is important to agriculture, commerce, customs and immigration, defense, ecology, fisheries, forestry, human health, and many other aspects of our society. Barcoding can meet these needs. Second, the Internet makes knowledge readily available and useful, and new approaches to managing information about organisms are capable of integrating molecular identification with other key information located in databases and web sites, so accelerating decision-making processes.

Names are an essential element in human activity and civilization. In choosing the things we name and in distinguishing things that have different names, we are encoding our perception of sameness and difference. When we create a name, we try to associate each name with a concept that has discrete, objective, and recognizable boundaries, like city limits or a color wheel that distinguishes blue from aquamarine. We communicate with each other using names for places, processes, objects, organisms, and so forth. The accuracy of any communication between two people depends upon our clear and shared understanding of the names. Without names and their associated definitions, we would not have organized societies.

Taxonomists are scientists who, as part of their profession, provide accurate names for identifying species. When species are named formally in scientific publications, there are standard formats that must be used to clarify the species concept associated with that name. In most cases, the species concept must be tied to a type specimen that is preserved in a herbarium, museum or culture collection as a permanent standard for comparison. In other cases, only an illustration is required and there is no physical embodiment of the species concept. In establishing and following standard procedures for naming species, taxonomists sit as one type of gatekeeper for human civilizations in that they provide the correct names for organisms so that all other humans can communicate quickly and accurately. Taxonomists are the custodians of biodiversity information.

“DNA barcodes” have recently been proposed as a way to assign biological specimens to their correct species, using a short standardized gene sequence. The barcode sequence can act as a proxy for a type specimen, allowing others to determine whether or not a specimen “matches” the species concept and can therefore bear the name of that species. Barcoding works by obtaining the gene sequence from either the type material or a “voucher” specimen that is known with certainty to belong to a particular species. These barcodes thereafter serve as references for the identification of unknown specimens. While DNA barcodes provide fast and accurate means for identifying

organisms, the barcode is only effective if the reference sequence is indisputably from the named species. That is, if the reference DNA barcode is mistakenly associated with the wrong species name, then the potential for significant and costly downstream errors will arise.

The purpose of the workshop was to address the problem of correctly associating reference DNA barcodes with species names for microorganisms known as protists (algae, fungi, protozoa). Today, there are approximately 200,000 described species of protists and approximately 92% of the currently tabulated species have a taxonomic description that is based upon an ink drawing (Appendices A, B). This opens up the problem of misidentification. This can be addressed in part using barcodes. Obviously, the ink drawing cannot yield DNA, and we need to resolve the problem of accurately linking the organism to a DNA barcode for most species. This general issue was divided into four areas to focus discussion during the workshop.

II. Summary of Workshop Conclusions

The workshop brought together scientists from around the world to address specific problems associated with protists, their names and their nomenclature and taxonomy (Appendices C, D). The participants reached consensus on a number of important issues and offered recommendations for the scientific community, governmental agencies, industry and other interested parties.

Participants in the workshop agreed that the published ink drawings associated with many species descriptions are inadequate and that epitypes and neotypes should be designated in these cases. They concluded that all new species descriptions should include DNA sequence data and designation of cryopreserved voucher specimens. Participants also agreed that a separate code of protistan nomenclature is not needed, though some specific changes to the botanical and zoological codes are needed. For example:

- The preservation of DNA should be required as part of the documentation of new species;
- Designation of epitype and neotype specimens should be allowed when holotypes are lost or cannot produce usable DNA;
- Cryopreserved material should be allowed as epitypes and neotypes;

The recommendations will be implemented in several ways. Recommended practices can begin immediately because they do not violate the Codes of taxonomic nomenclature. Codification of some recommendations to the International Codes of Nomenclature requires formal procedures that will take years to implement, nevertheless this process should begin immediately. The selection of specific cultures to represent biological type material is not a trivial matter, and this key step should begin immediately. However, it is likely that the process will be laborious and therefore financial support should be sought to ensure that this step is implemented. Similarly, the informatics work is laborious and financial support should also be sought to ensure implementation of this critical part. There is every reason to believe that protistan

barcodes can be tied to species names by using cultures as the DNA source. Informatics can unite critical information so that it is quickly and easily accessible to a wide range of users. Within a few years, governmental agencies, industry and research scientists will have a quick and reliable means for identifying protists.

III. Background Information on Protists

Protists are an assemblage of mostly microscopic eukaryotic organisms commonly referred to as algae, aquatic fungi, and protozoa. Many protists are pests or pathogens (e.g., malaria, rusts) and others provide critical ecosystem services (e.g., photosynthesis and oxygen release, decomposition of organic matter). Despite approximately 200 years of study, protists are poorly known compared to animals and plants. Even the estimated numbers of species vary substantially. A summary of protists was reported by Corliss (1982) and is reproduced here as Appendix A. As part of this workshop, a second partial tabulation of species was assembled to specifically address the number of species where DNA may be obtained from the taxonomic type material (Appendix B).

Protists cause numerous diseases in humans, animals, and plants. Malaria kills more humans each year than any other disease, accident or war. Malaria is caused by apicomplexans named *Plasmodium* (i.e., *Plasmodium falciparum*, *Plasmodium vivax*, etc.). A related group of apicomplexans named *Toxoplasma* cause toxoplasmosis and is implicated in other human diseases, such as Alzheimer's disease. Other common protistan diseases include amoebic dysentery (*Entamoeba*), African sleeping sickness and Chagas' disease (*Trypanosoma*), Leishmaniasis (*Leishmania*) and various intracellular parasites (e.g., *Brachiola*, *Encephalitozoon*, *Microsporidium*, *Septata*). Animal diseases caused by protists include coccidiosis (in dogs, cats, poultry; e.g., *Isospora*), hexamitiasis (=infectious catarrhal enteritis; *Hexamitis*) in turkeys, blackhead disease (*Histomonas*) in poultry, and tail rot in aquarium fishes (*Saprolegnia*). The famous Irish potato famine was caused by *Phytophthora*, downy mildew of grapes by *Plasmopara*, molds on grass and leaves by *Peronospora*, white rust on pine trees by *Albugo*, and root rot of peas by *Aphanomyces*.

Protists are more often nondestructive organisms that play key roles in ecosystems. Algae and cyanobacteria (=blue-green algae) are the near exclusive source of photosynthesis in the oceans – there are no grasslands or forests in seawater. The oceans cover 71% of the earth's surface. The marine algae and cyanobacteria account for approximately 50% of the global photosynthesis, and thus the oxygen in every other breath that we breathe comes from a marine microbe. Directly or indirectly they feed almost all marine animals (e.g., clams, shrimp, lobsters, fishes, sea birds, whales). Protists are also crucial for the process of recycling compounds so that they can be re-used in the ecosystems. Furthermore, petroleum (crude oil, natural gas) is derived from marine algae, whereas coal comes from plants.

In naming and describing species of protists, taxonomists follow either the International Code of Botanical Nomenclature (algae, aquatic fungi) or the International

Code of Zoological Nomenclature (for protozoa). The botanical code allows an illustration if and only if a biological specimen cannot be preserved, while the zoological code always requires a type specimen. In addition to this variation of practice within protistan taxonomy, there are a number of other issues that complicate the relationship between protist names and their associated species concepts:

- Historically, microorganisms were not preserved, and new species were often described using drawings made while the organism was observed on temporary wet-mount microscope slide preparations.
- Frequently, populations, not individuals, were preserved when permanent microscope slides were prepared. In some cases (e.g., diatoms), the type was distinguished by marking the coverslip with a permanent circle around the specific cell(s).
- DNA cannot be recovered from preserved slide specimens, in general. This is especially true for diatoms because their organic matter was digested so that frustule markings were visible.
- Ink illustrations of protists varied widely in their quality, and in many cases they are inadequate representations of a species concept. A cryptic species complex may be revealed by modern electron microscopy or DNA sequence analysis, however, the historic illustrations cannot be assigned to a specific species of the complex.
- Many protist species are culturable and are being perpetuated in culture collections. However, nomenclatural codes do not accept actively growing cultures as type specimens because they can accumulate genetic and morphological changes over time.
- Cultured lineages can be preserved cryogenically, allowing for DNA extraction while arresting genetic and morphological changes.

Participants stressed the need to use the most modern approaches to protist taxonomy, include DNA barcoding, web-based databases, and an electronic registry of names.

IV. The Workshop

The workshop was held at the Portland Harbor Hotel on November 6 and 7, 2006. The participants are listed in Appendix C. The agenda for the workshop is presented in Appendix D. Formal talks for the meeting are available on the Consortium for the Barcode of Life (CBOL) website:

<http://barcoding.si.edu/SloanProtistMeetingAgenda2006.htm>

The workshop proceeded rather quickly because there was immediate consensus on several important topics. It was agreed that molecular sequence data (DNA), which are currently being used to identify species, will increasingly become the primary means for identification of microorganisms. It was agreed that barcodes provide a critical step toward the goal of DNA identification for protists, but everyone agreed that barcodes cannot be used to establish evolutionary relationships (phylogenetic trees) and barcodes by themselves cannot be used to describe new species. We also agreed that it was

fruitless to discuss the gene choice for the barcode sequence (e.g., CO1, ITS) because there is currently insufficient scientific data to evaluate the choice of the best gene and because technology is changing so rapidly that within a few years much more detailed and sophisticated barcodes may exist. Consequently, *the workshop chose to be visionary rather than reactionary, and subsequent discussions were aimed at finding solutions that would not only work today but would work well in future years.*

Barcoding itself was not discussed in detail during the workshop, but several important facts should be presented. The Canadian Barcode of Life Network, through several different scientific laboratories, is testing several different barcode sequences (e.g., CO1, ITS, SSU rRNA). Many laboratories are using traditional sequencing, although 454 sequencing is gaining popularity. New, radically different technologies will likely appear in the next few months or years, and existing problems will likely be solved in the near future. Early results suggest that a successful barcode marker will be selected within the coming months. Primer development is challenging, but steady progress is being made. Efforts are currently underway to examine the breadth and depth of barcode markers. The breadth is measured by using the same barcode across the wide range of protistan diversity (as well as animals, fungi, plants). The depth is measured by obtaining the identical barcode sequence from numerous representatives of a single species.

V. The Problems

Problem 1

How do we connect DNA from a culture to the correct name?

Officially, scientific names of organisms are regulated by codes of nomenclature. The International Code of Botanical Nomenclature and the International Code of Zoological Nomenclature regulate protistan nomenclature. The reference material should be a specimen, but the codes allow for the use of ink drawings or photographs under some circumstances. Barcodes require DNA, and DNA cannot be obtained from drawings and photographs. DNA is also unavailable from some biological specimens, such as permanent microscope slides, old herbarium specimens, and specimens embedded in plastic. Therefore, the DNA for microbial protists will be obtained almost exclusively from expertly curated living organisms held in culture collections. Worldwide, there are approximately 200 culture collections that contain protists. For many species, numerous cultures exist.

Recommendations for Problem 1

How do we connect DNA from a culture to the correct name?

There was extensive discussion about the botanical and zoological codes of nomenclature, and several issues were raised. The following conclusions and suggestions were agreed upon:

- (1) The current codes have problems but we should work to change the existing codes, and there should be no effort to establish a separate protistan code of nomenclature.
- (2) The current codes should be modified so that they require some type of DNA sequence information for the description of all new species: no longer should a microbial

organism be described using only an ink drawing, an electron micrograph or other type of nonbiological representation.

(3) The use of epitype and neotype (that is, the establishment of unambiguous reference material for the species) options, should be employed when existing descriptions are ambiguous (e.g., when the holotype and other type material cannot yield DNA).

(4) Designation of new epitypes and neotypes for microbes should be based upon a (cryopreserved) culture when available. The codes should be modified, where necessary, to make this easier.

Problem 2

Which of the many cultures should be chosen to be the “official” culture upon which the barcode becomes anchored?

Cultures used to describe species or cultures from the original collection site (i.e., type locality) are preferred. The type locality can be revisited in some cases, but in other cases that locality may be covered with houses, polluted and no longer able to support growth of that species. In the case of marine protists, the water body that contained the population on which a new species name was based may have moved owing to changes in circulation pattern. The notion of a type locality is therefore difficult to apply in the oceans. When either the type culture is lacking or the type locality is no longer accessible, another culture source must be chosen. A set of guidelines for selecting that “official” culture should be established.

Recommendations for Problem 2

Which of the many cultures should be chosen to be the “official” culture upon which the barcode becomes anchored?

Participants agreed unanimously that, when DNA-yielding type material is absent, a single, specific culture strain should be tied to the name. It was recognized that cultures are not always available, especially for parasitic organisms that require a living host organism. It was agreed that some current cultures may be misidentified and it should not be assumed that every culture is properly identified. It was agreed that some cultures, identified as a single species based upon morphology, might represent a cryptic species complex. It was also unanimously agreed that if a specific culture is designated as type material, it must agree with the original diagnosis; if the organism in culture differs from the original diagnosis, then it cannot be used as an epitype or neotype. These points are carefully detailed in the botanical and zoological codes and cannot be ignored. With regard to the choice of a culture to anchor an existing scientific name, the following recommendations were made:

(1) The choice of a specific, single culture strain that will be the source of epitypification or neotypification should adhere to the following guidelines:

- (a) If an actual type culture exists, this specific strain must be used. If not available, then
- (b) An authentic culture strain must be used if it exists. If not available, then
- (c) A culture strain from the type locality must be used if it exists. If not available, then

- (d) A culture strain from the same continent and habitat type must be used if it exists. If not available, then
- (e) A culture strain from the same habitat type must be used if it exists. If not available, then
- (f) A culture strain that best represents the species must be used.

(2) Wherever and whenever possible, new type material should include a cryopreserved culture strain.

(3) The name of the species is tied to the specific culture (via epitypification or neotypification). If other cultures bearing that name are believed to belong to a different species, then they must be given different names (i.e., other existing names or described as a new species).

(4) Scientific debate on the separation or combination of species must follow the codes. That is, splitter and lumper discussions will be taxonomy as usual, but the names will be anchored to biological material (e.g., a specific culture) so that biological investigation can be part of the discussions.

Problem 3

How do we coordinate the names, the cultures, the barcode sequences and important information for each species?

Modern identification must not only be rapid and accurate, but it must also be easy to use and accessible worldwide. It is obvious that the information will be managed on electronic databases and transmitted by the Internet. However, assembling databases, linkages and user-friendly portals requires considerable effort. Although not part of the original workshop topic, it quickly became apparent that management of information with informatics tools was crucial to successful implementation of other workshop recommendations.

Recommendations for Problem 3

How do we coordinate the names, the cultures, the barcode sequences and important information for each species?

Workshop participants representing CBOL, the Canadian Barcode of Life Network, the National Center for Biotechnology (NCBI, parent organization of GenBank), micro*scope, Encyclopedia of Life, International Census of Marine Microbes, and the Marine Microbes Forum discussed the application of new informatics solutions to the dynamic integration of databases and web sites distributed across the internet. The bar-coding technology should be able to access this information and bring relevant information to the attention of the users. Barcode sequences in genomic repositories (e.g., DDBJ, EMBL, GenBank) should state that the barcode comes from the type material for that species or from other material that is believed to be representative of that species. There should be voucher specimens for all barcodes. For microbial organisms, the culture can be a voucher but electronic images of the organism from the culture should also be available. There should be biogeographical, ecological, and environmental data associated with the organism, its barcode and its culture. Therefore, informatics is crucial

and must be part of the action plans. The following are recommendations from the workshop:

(1) A registry of names should be established. As part of the registry, a nomenclator for protists should be included to find, validate, and annotate names. No such nomenclator exists at this time, but the taxonomic structure used by micro*scope is, with appropriate investment, well positioned to fulfill this role.

(2) The registry should be linked to biogeographical, ecological, and environmental data. The ICoMM MICROBIS database was developed to meet this need; and coupled with the cyberinfrastructure associated with the EoL project, will, with appropriate investment, be able to interface with a wide variety of electronic databases and web sites.

(3) The cyberinfrastructure should be designed to allow dynamic internet links between culture collections, genomic repositories, taxonomic web sites, and so forth.

(4) An on-line index of all cultures should be prepared, ideally in a dynamic structure that is automatically, or at least easily, updated.

Problem 4

How do we implement the barcode/culture/name process?

The Canadian Barcode of Life Network has several groups at different Canadian universities who are currently evaluating barcode markers (e.g., CO1, ITS, SSU rRNA). The Consortium for the Barcode of Life (CBOL) has also organized an international workshop on the selection of an optimal barcode. The workshop is being supported by a grant from the Sloan Foundation to Rutgers University and will take place on 14-15 May 2007 in Front Royal, VA. The successful marker(s) would work across the protists (breadth) and would discriminate closely related species. These preliminary tests will be completed within a few months and the data will determine the appropriate barcode gene for protists. The workshop participants from culture collections held a break-out group to discuss implementation of the next steps in this process.

Recommendations for Problem4

How do we implement the barcode/culture/name process?

Participants representing culture collections, and possibly other culture collections not represented at the workshop, were thrown a “1,000 strain challenge” from the Canadian Barcode of Life Network. The culture collections agreed to provide cultures at no cost to the Canadian Network, while the Network, by reciprocation, agreed to provide barcodes at no cost to the culture collections.

The list should include organisms that (a) cover the breadth across all protistan groups, and ideally a single barcode gene would be applicable, (b) typically include species where 5-10 culture strains are available to provide depth, but (c) include species with less than 5 culture strains when it is necessary to achieve taxonomic breadth. The list would be restricted to microscopic protists: seaweeds would not be included. The list of 1,000 strains would not emphasize groups that are being intensely studied by the Network (e.g., ciliates, diatoms, dinoflagellates).

Workshop participants representing GenBank and EMBL also discussed implementation from the standpoint of a genetic repository. These institutions have established procedures for identifying barcodes and they will continue to adjust these procedures with the changing needs of the scientific community.

Appendix A. List of Protistan Groups and the Number of Species for Each Group. From Corliss, J.O. 1982. J. Proto. 29: 499

Name of Group (with succinct description or comment)	No. of Species
Acrasids (cellular slime molds)	26
Actinopod amoebae (axopodiated forms with internal skeletons: heliozoa, acantharians, radiolarians) 40% fossil	7,000
Bacillariophytes (diatoms: I've excluded probable doubtful - but apparently recorded - forms, with which total number of species in this algal group might reach 50-60,000, or even 100,000!) 60% fossil	25,000
Charophytes (stoneworts s.l. and relatives 75% fossil)	400
Chloromonads (raphidophyte algae)	27
Chlorophytes (green algae s.s. + 800 phytomonad volvocines)	3,200
Choanoflagellates (craspedomonad algae: collared cells)	140
Chryomonads (golden algae s.s., including silicoflagellates, etc. and also bicosoecids) 25% fossil	850
Chytridiomycetes (phycomycetes group with flagellated zoospores)	900
Ciliates (all heterokaryotes, with loss of <i>Stephanopogon</i>)	7,500
Cryptomonads (assorted forms, unclear relationships)	200
Dinoflagellates (pyrrhophytes s.l.: diverse forms + ebrriids + some acritarchs; the "mesokaryotes") 50% fossil	4,200
Euglenoids (euglenophytes + 550 kinetoplastids + <i>Stephanopogon</i> !)	1,600
Eumycetozoa (plasmodial myxomycetes and relatives)	550
Eustigmatonomads (unicellular xanthophyte group)	11
Foraminiferans (reticulopodiated forms with tests) 80% fossil	37,500
Haplosporidians (ascetosporans, excluding some doubtful forms)	30
Haptomonads (prymnesiomonads, mainly coccolithophores) 70% fossil	1,500
Heterochlorids (unicellular xanthophyte group)	15
Hyphochytriomycetes (phycomycetes group with flagellated zoospores)	25

Labyrinthulids (net slime molds)	36
Metamonads (“higher” zooflagellates; various taxa + parabasalids)	2,200
Microsporidia (spore with polar cap and filament)	800
Myxozoa (myxosporidians; valved, multicellular spores)	875
Oomycetes (water molds and relatives with flagellated zoospores: many species doubtful)	800
Opalinids (paraflagellates: only half may be valid species)	400
Pelobionts (karyoblasteans: 20 spp., yet all = <i>Pelomyxa palustris</i> ?)	1
Phaeophytes (brown algae: seaweed, kelp)	1,600
Plasmodiophorids (endoparasitic slime molds)	36
Prasinomonads (primitive green algae, some acritarchs) 25% fossil	350
Proteromonads (uncertain group, now greatly reduced in size)	10
Rhizopod amoebae (naked + shelled forms + amoeboflagellates + 36 xenophyophores, but excluding ca. 2,000 doubtful testacea) 15% fossil	2,500
Rhodophytes (red algae: seaweeds; remote ancestry) 15% fossil	5,000
Sporozoa (apicomplexans: gregarines, coccidians, haemosporidians, piroplasms)	4,800
Xanthophytes (yellow-green algae s.s.)	650
Zygnematophytes (conjugating algae, with >4,000 desmids: plus perhaps an additional 5,000 ⁺ desmids?)	4,800
Total Number of Protistan Species	115,532

Appendix B. A list of taxonomic groups, the number of valid species for each group, the estimated total number of species (including undescribed species), the number of species where the nomenclatural type material is an ink drawing or similar printed image, the number of species where the type material consists of biological material, the number of species based upon authentic cultures, and the number of species that have type material containing DNA that is accessible using modern techniques. Data assembled by workshop participants.

Taxon	# valid spp.	Estimated total # spp.	Type ink, EM etc.	Type Biological	Authentic Culture	Type with DNA
Aphelidea	11		8	NA	1	3
Apicomplexa	4296	4,541,500	375	4005	100	100
Apusomonadida	11	20	10	NA	1	1
Bicosoecida	47	100	10	NA	3	4
Cercomonadida	72	150	59	NA	6	13
Chlorarachniophyta	9		2	1	4	0
Choanoflagellida	161	200	140	NA	13	22
Chrysophyceae	700	2000	675	25	5	0
Ciliates	9000	30,000	8000	1500	99	99
Conjugating Green algae	7000	21,000	6000	1000	0	0
Cryptomonas	126	150	111	1	14	14
Diatoms	100,000	1,000,000	99,988	0	12	12
Dinoflagellates	2000	3500	1970	??	30	0
Euglenids	1400	1700	1300	60	15	12
Foraminifera	37,500	4000	33,500	4000	0	0
Haptophyta	300	750	290	10	30	0
Labyrinthulomycota	50	500	40	20	10	12
Oomycetes	2000	4000	500	250	250	0
Phaeophyceae	1800	2000	0	1800	0	500
Phaeothamniophyceae	28	100	20	8	0	0
Pinguicophyceae	5	15	1	4	4	0
Schizocaldiophyceae	1	5	1	1	1	1
Synurophyceae	287	175	287	0	0	0
Thaumatomonadida	34	50	31	NA	2	3

subtotal	166,838	5,611,915	153,318	12,685	600	796
percent of valid species	100.00%	NA	91.90%	7.60%	0.36%	0.48%

Appendix C. Names of participants at the Alfred P. Sloan Foundation Workshop, their institutions and their e-mail addresses. [Note: This table should probably be structured differently for the final report (e.g., no e-mail address; with address and country columns). However, this is a quick table for now.

Name	Institution	e-mail address
Andersen, Robert	Bigelow Laboratory, USA	randersen@bigelow.org
Beakes, Gordon	Newcastle Univ., UK	g.w.beakes@ncl.ac.uk
Blackburn, Susan	CSIRO, Australia	susan.blackburn@csiro.au
Brand, Jerry	Univ. Texas, USA	Jbrand@mail.utexas.edu
Cochrane, Guy	EMBL-EBI	cochrane@ebi.ac.uk
Draisma, Stefano	Leiden Univ., Netherlands	draisma@nhn.leidenuniv.nl
Duszynski, Donald	Univ. Arizona, USA	eimeria@unm.edu
Edvardsen, Bente	Univ. Oslo, Norway	bente.edvardsen@bio.uio.no
Federhen, Scot	GenBank, USA	federhen@ncbi.nlm.nih.gov
Friedl, Thomas	Univ. Goettingen, Germany	tfriedl@gwdg.de
Hajibibael, Mehrdad	Univ. Guelph, Canada	mhajibab@uoguelph.ca
Hoef-Emden, Kerstin	Cologne Univ., Germany	aeb25@uni-koeln.de
Ishida, Ken	Tsukuba Univ., Japan	ishidak@kenroku.kanazawa-u.ac.jp
Kaczmarska, Irena	Mt. Allison Univ., Canada	iehrman@mta.ca
Karpov, Sergey	St. Petersburg, Russia	sakarpov4@yandex.ru
Kawachi, Masanobu	NIES, Japan	kawachi@nies.go.jp
Kawai, Hiroshi	Kobe Univ., Japan	kawai@kobe-u.ac.jp
Küpper, Frithjof	CCAP, U.K.	Frithjof.Kuepper@sams.ac.uk
Leander, Brian	Univ. Br. Col., Canada	bleander@interchange.ubc.ca
Levesque, Andre	Agriculture Canada, Ottawa	LevesqueCA@AGR.GC.CA
Lewis, Louise	Univ. Connecticut, USA	louise.lewis@uconn.edu
Lynn, Denis	Univ. Guelph, Canada	ddr@uoguelph.ca
Moestrup, Øjvind	Univ. Copenhagen, Denmark	MOESTRUP@bot.ku.dk
O'Kelly, Charles	Bigelow Laboratory, USA	cokelly@bigelow.org
Patterson, David	M.B.L., USA	dpatterson@mbl.edu
Phang, Siew Moi	Univ. Malaysia, Malaysia	phang@um.edu.my
Porter, David	Univ. Georgia, USA	porter@plantbio.uga.edu
Potapova, Marina	Phil. Acad. Nat. Sci. , USA	potapova@acnatsci.org
Probert, Ian	Univ. Caen, France	algobank@unicaen.fr
Pröschold, Thomas	CCAP, U.K.	Thomas.Proeschold@sams.ac.uk
Saunders, Gary	Univ. New Brunswick, Canada	gws@unb.ca
Schindel, David	CBOL, , USA	schindeld@si.edu
Silva, Paul	Univ. Calif.-Berkeley, USA	psilva@uclink4.berkeley.edu
Spiegel, Fred	Univ. Arkansas, USA	fspiegel@uark.edu
Stern, Rowena	Univ. Br. Col., Canada	rowenas@interchange.ubc.ca
Theriot, Edward	Univ. Texas, USA	etheriot@mail.utexas.edu
Vargas, Colomban de	Roscoff Biol. Lab, France	vargas@sb-roscoff.fr

Vaulot, Daniel
Williamson, Phil

Roscoff Biol. Lab, France
NERC, U.K.

vaulot@sb-roscoff.fr
p.williamson@uea.ac.uk

Appendix D. Agenda for the Alfred P. Sloan Foundation Workshop.

Monday, November 6, 2006

- 9:00 – Welcome – Robert Andersen and Frithjof Küpper
- Morning Discussion Chair – Kerstin Hoef-Emden
- 9:15 – Introduction of Participants
- 9:30 – Robert Andersen – Overview of the Workshop and the Problem
- 9:50 – Frithjof Küpper – Role of Barcoding at Culture Collections
- 10:10 – Mehrdad Hajibabaei – The Canadian Barcode of Life Network (or the Barcode of Life Initiative)
- 10:30 – Coffee Break
- 11:00 – David Patterson – Barcodes and Zoocodes
- 11:30 – Paul Silva – Barcodes and the Botanical Code of Nomenclature
- 12:00 – Lunch
- Afternoon Discussion Chair – Brian Leander
- 1:30 – Open discussion on problems associated with establishing the correct linkage between barcodes and taxonomic types, authentic cultures, etc.
- 3:30 – Coffee Break
- 4:00 – Open discussion on problems associated with misidentified or mislabeled cultures that are revealed by barcodes
- 5:00 – Close of First Day Session
- 7:00 – Evening Dinner

Tuesday, November 7, 2006

- 9:00 – Robert Andersen – Announcements (if any)
- 9:10 – Division of participants into two working groups
 - a. Taxonomic group (Chair, Øjvind Moestrup; Recorder, Denis Lynn)
 - b. Culture collection group (Chair, Jerry Brand; Recorder, Thomas Friedl)
- 9:15 – Charge for each working group
- 10:30 – Coffee Break
- 11:00 – Continued working group discussions
- 12:30 – Lunch
- Afternoon Discussion - Robert Andersen and Frithjof Küpper (chairs)
- 1:30 – Recommendations from Taxonomic Group
- 1:45 – Recommendations from Culture Collection Group
- 2:00 – Informatics discussion (David Patterson, Daniel Vaultot co-chairs)
- 3:30 – Coffee Break
- 4:00 – Summary and Concluding Remarks- Andersen, Küpper (chairs)
- 5:00 – Close of Workshop